UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DC

AD-A252 181

DTIC ELECTE JUN 3 0 1992 S A D

1a. REPORT SECURITY CLASSIFICATION	
2a. SECURITY CLASSIFICATION AUTHORITY	...AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Essex Corporation; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: USAF School of Aerospace Medicine
6c. ADDRESS (City, State, and ZIP Code): 1040 Woodcock Road, Suite 227 Orlando, FL 32803	7b. ADDRESS (City, State, and ZIP Code): Human Systems Division Brooks Air Force Base, OH 78235-5301
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: USAF School of Aerospace Med.; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F33615-88-C-0640
8c. ADDRESS (City, State, and ZIP Code): Brooks Air Force Base, TX 78235-5301	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Enhancing Cognitive Performance Through Biocybernetic Feedback: An Individual Differences Approach

12. PERSONAL AUTHOR(S)
Kennedy, Robert S., Fowlkes, Jennifer E., and Smith, Martin G.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 11/88 TO 6/89	88/06/14	28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Workload, vigilance, biocybernetics, eye moyement metrics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

For the present investigation, we systematically varied a workload task over a range of difficulties and sessions while eye movement elements and performance served as dependent variables. The objective was to determine whether task demands which are inherent in the stimulus (e.g., number of channels monitored; time on task) covaried with characteristics of the dependent variables. Although the approach was empirical, the elements selected for study followed from theory or findings reported previously in the scientific literature. The immediate goal of this work was to surface a series of objective indicators (both performance and bioelectric events) which vary systematically with workload (both operator dependent and operator independent). The intermediate goal is to develop a battery of measures and automated scoring algorithms to index the different forms of workload effects and operator state. The long-term goal of this work is the

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. William F. Storm	(512) 536-3464	USAFSAM/VNB

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

19. Abstract (continued)

development of a biocybernetic device. Such a device would employ these suitable biomedical measures, analyze them in real time, and feed the information back to the machine (or operator) in order to modify systems performance.

For the present Phase I report, three separate efforts were conducted. The first, a pilot study, employed five subjects to whom performance tests were administered while the feasibility of the eye movement recording techniques was explored and to gain experience with the scoring procedures. Then, in the main experiment, 15 subjects were exposed to three different levels of workload difficulty over six sessions (each workload condition repeated twice). Performance (percent correct) on the tasks and eye movement metrics (blink rate, frequency, amplitude, and velocity of saccades) were recorded throughout. In addition to the task difficulty, "time on task" was examined as a dependent variable on bioelectric and performance metrics. In a third study, longer sessions were employed to gain experience with the effects of sustained performance on the eye movement and performance metrics.

In the main experiment, it was demonstrated that group performance (percent correct) on the tone counting task varied with task difficulty ($p < .01$), suggesting that the workload had been successfully manipulated. Of the eye movement measures, acceleration of eye movements (or the slope of the regression line relating velocity of saccades to amplitude) bore a strong relationship to task difficulty, becoming steeper (faster eye movements) for 80% of the subjects ($p < .01$) when the high task loading condition was compared to the lower. Eye movement frequency and eye blinks did not appear to be different in the two task loadings ($p = .40$). For the other eye movement measures explored: (a) amplitude of saccades tended to increase in the high task load condition for the majority of the subjects (61%) but this difference was not significant ($p < .15$), and (b) aggregated eye movement velocities under the high workload were generally greater in the high workload condition for the majority of the subjects (73%; $p < .02$).

While the relationship between time on task and performance followed predicted relationships, the size of the effect was small (4%) and not statistically significant ($p < .20$). Although performance may not have degraded under the conditions as were measured, bioelectric measures nonetheless were predictive of time on task: (a) frequency of eye movements declined for 73% ($p < .037$) of the subjects; and (b) acceleration of eye movements increased for 66% ($p = .098$) suggesting that both metrics hold promise as prospective measures of sustained performance.

Finally, as part of the present effort, we also have developed considerable customized computer techniques and algorithms for rapid scoring and analysis of these key eye movement metrics. Presently, 60% of all scoring for the eye movement activity system is fully automatic (i.e., eye blink and saccade metrics) and the remainder is conducted interactively using an experimenter. This reduces by 90% the scoring process for the experimenter and was seen as a necessary intermediate step in the development of a fully automated scoring system. We believe there are no technical impediments to improving the automatic scoring techniques so that analyses could be conducted in near real time with more sophisticated equipment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. EXECUTIVE SUMMARY

Workload measurement involves an attempt to characterize conditions under which task demands can or cannot be met by the performer. In the early days of aviation, the machine was often the limiting factor in system performance. Today, overstressed aviators may be more common than overstressed aircraft. New systems permit more information to be presented in real- or fast-time to the operator than he or she can handle efficiently. In modern weapon systems the human sensory systems are overloaded and the problem is not likely to abate. While military equipment is becoming more complex, the proportion of the available pool of personnel which possesses the mental capabilities to accommodate these emerging systems is becoming more limited (Tice, 1986; Merriman & Chatelier, 1981).

There are many ways in which workload may be studied. The investigations presented here presume that biological events may be predictive of the task demands on the operator as well as the individual's personal attention to that work. For example, we know there may be little or no deterioration in operational performance until the point of failure is closely approached, but perhaps sensitive biological measures of workload could provide premonitory signs of impending failure. However, to investigate such a set of relations, we believe that "workload" needs to be measured. But before it can be measured, it needs to be defined. For this purpose we have adopted a model presented in a previously conducted project sponsored by the U.S. Air Force. In that work, we distinguished between workload measures which seek to characterize the stimulus (we call them operator-independent measures) and those which may be different for different subjects (we call them operator dependent). We believe it is possible to have objective measures (in this case bioelectric events and performance) which can index both constructs. We recognize that logically there can also be subjective measures of both of these constructs, but such instruments are not a part of this work.

For the present investigation, we systematically varied a workload task over a range of difficulties and sessions while eye movement elements and performance served as dependent variables. The objective was to determine whether task demands which are inherent in the stimulus (e.g., number of channels monitored; time on task) covaried with characteristics of the dependent variables. Although the approach was empirical, the elements selected for study followed from theory or findings reported previously in the scientific literature. The immediate goal of this work was to surface a series of objective indicators (both performance and bioelectric events) which vary systematically with workload (both operator dependent and operator independent). The intermediate goal would be to develop a battery of measures and automated scoring algorithms to index the different forms of workload effects and operator state. The long-term goal of this work is development of a biocybernetic device. Such a device would employ these suitable biomedical measures, analyze them in real time, and feed the information back to the machine (or operator) in order to modify systems performance.

For the present Phase I report, three separate efforts were conducted. The first, a pilot study, employed five subjects to whom performance tests were administered while the feasibility of the eye movement recording techniques was explored and to gain experience with the scoring procedures. Then, in the main experiment, 15 subjects were exposed to three different levels of workload difficulty over six sessions (each workload condition repeated twice). Performance (percent correct) on the tasks and eye movement metrics (blink rate, frequency, amplitude, and velocity of saccades) were recorded throughout. In addition to the task difficulty, "time on task" was examined as a dependent variable on bioelectric and performance metrics. In a third study, longer sessions were employed to gain experience with the effects of sustained performance on the eye movement and performance metrics.

In the main study, it was demonstrated that group performance (percent correct) on the tone counting task varied with task difficulty ($p < .01$), suggesting that the workload had been successfully manipulated. Of the eye movement measures, acceleration of eye movements (or the slope of the regression line relating velocity of saccades to amplitude) bore a strong relationship to task difficulty, becoming steeper (faster eye movements) for 80% of the subjects ($p < .01$) when the high task loading condition was compared to the lower. Eye movement frequency and eye blinks did not appear to be different in the two task loadings ($p = .40$); however, other eye movement measures were explored: (a) amplitude, the SLIT metric of previous AFOSR work, tended to show greater amplitude in the high task load condition for more subjects but this difference was not significant (61%; $p < .15$); also, and (b) aggregated eye movement velocities under the high workload were generally greater for more subjects (73%; $p < .02$).

While the relationship between time on task and performance followed predicted relationships, the size of the effect was small (4%) and not statistically significant ($p < .20$). Although performance may not have degraded under the conditions as were measured, bioelectric measures nonetheless were predictive of time on task: (a) frequency of eye movements declined for 73% ($p < .037$) of the subjects; and (b) acceleration of eye movements increased for 66% ($p < .098$) suggesting that both metrics hold promise as prospective measures of sustained performance.

In summary, a proof-of-concept effort was conducted in order to determine whether characteristics of workload and arousal could be examined behaviorally (percent correct) and biomedically (eye movement measures). It was found that a performance test as an independent variable served as an exteroceptive stimulus (task control of performance) and certain eye movement characteristics were surfaced which covaried with task complexity. Other related eye movement metrics also showed promise. The strength of relationship for an interoceptive stimulus (viz., time on task) was weaker, but eye movement metrics were surfaced which appeared to change over short sustained periods. Longer periods of exposure to performance tests during eye movement recording were demonstrated successfully in a few subjects in order to verify the feasibility of this procedure for sustained performance experimentation.

Finally, as part of the present effort, we also have developed considerable customized computer techniques and algorithms for rapid scoring and analysis of these key eye movement metrics. Presently, 60% of all scoring for the eye movement activity system is fully automatic (i.e., eye blink and saccade metrics) and the remainder is conducted interactively using an experimenter. This reduces by 90% the scoring process for the experimenter and was seen as a necessary intermediate step in the development of a fully automated scoring system. We believe there are no technical impediments to improving the automatic scoring techniques so that analyses could be conducted in near real time with more sophisticated equipment. However, we believe it is first necessary to determine what biomedical elements to score. To this end, we believe this system can provide an important adjunct to our Automated Performance Test System (APTS) and can have broad implications for behavioral performance assessment. For example, all of the APTS tests are automatically scored and would permit the biomedical study of a fully researched, factorially diverse cognitive and motor battery of performances, coincidentally with biomedical relations like eye movements to which they may be differentially related.

II. INTRODUCTION

In the early days of aviation, the machine was often the limiting factor in system performance. Today, overstressed aviators may be more common than overstressed aircraft, and recently reported Stealth crashes (Orlando Sentinel, June 6, 1989) imply that fatigue following heavy schedules were contributing factors. New systems permit more information to be presented in real- or fast-time to the operator than he or she can handle efficiently. Descriptive terms like "getting behind the system," "information overload," and "noisy" abound in design conferences, and, poignantly, there are "declutter switches" in heads-up displays. In modern weapon systems the human sensory systems are often so bombarded with stimuli that inferences from displayed information are time constrained and performance suffers. These issues are not limited to military jobs. Civilian air traffic controllers are often retired for job-related workload stress (Hale, Williams, Smith, & Melton, 1971), and astronauts experience "time compression" in connection with their high workload periods upon reentry (Schmitt & Reid, 1985). As military equipment has been made more complex, the proportion of the available pool of personnel which possesses the mental capabilities to accommodate these emerging systems is becoming more limited (Tice, 1986; Merriman & Chatelier, 1981).

Problems caused by increasing task demands do not come as any surprise to those who grew up in the field of human factors engineering. Indeed, they were eloquently predicted by Chapanis, Garner, Morgan, and Sanford (1947) in one of the earliest works on systems research where they said, "...Let us look ahead about 50 years [it's only 42!] and imagine what kind of problem the air traffic control officer at LaGuardia Field will have when a heavy fog settles over the place and 100 planes converge on the field from Mexico, London, Paris, San Francisco, and Albuquerque. How are we going to get all that information to him [sic!]. How can we present this information to him so that he can see the total picture? There are, after all, only a few channels that we can use in getting this information into his brain" (p. 240). Relatedly, and more recently, in a Congressional hearing, several prominent human factors experts from different fields, inside and outside the Department of Defense, were invited to offer their opinions as to the cause of the downing of the Iran Air Flight 655. Granted that an incorrect decision had been made from displayed information, the one issue about which the experts appeared concordant was that the incident was due to the combined effects of the high workload and combat stresses which were not accommodated in the human factors design of the tactical information displays for the AEGIS weapon system (House Armed Services Committee, 1988).

The idea which prompted the present research was that biological events may be reflective of the task demands of the work as well as the operator's attention to the task. We know that there may be little or no deterioration in operational performance until the point of failure is closely approached (Gopher & Donchin, 1986; Schmidt, 1978), but perhaps sensitive biological measures of workload could provide premonitory signs of impending failure. There are two technical developments which must be accomplished to produce a workable system. One is traditionally a human factors effort, and in the case of bioelectric events, includes several aspects of the biomedical sciences and technology. The second is the engineering development, including software and hardware which must accompany such an enterprise. In this work we have set out to accomplish both, but the emphasis has always been on the former rather than the latter.

There are many applications for validated biocybernetic workload indices. To the extent that such workload measures also follow task demands, they could be employed to index workload characteristics of military systems during various stages in the human factors engineering design and subsequent evaluation. It is generally accepted that workload is a function of task, operator capabilities, and sequential characteristics (e.g., practice with task, previous workload history, etc.). Moreover, a general consensus is that workload is multidimensional and that the researcher should have a battery of workload metrics at his or her disposal (Bateman, 1979; Crabtree, Bateman, & Acton, 1984; Eggemeier, 1980; Eggemeier, Shingledecker, & Crabtree, 1985; Frazier & Crombie, 1982; Gopher & Donchin, 1986). Researchers are determining whether there are reliable indicants of individual differences in information processing skills (Benel, Coles, & Benel, 1979; Damos, 1984a,b,c; Damos & Smist, 1981; Wickens, Mountford, & Schreiner, 1981). By assessing workload's impact on individuals for tasks of constant difficulty, workload indices can be used to match systems according to the reliable individual differences of the user population. Then better decisions can be made selecting between design tradeoffs (e.g., customize for individual use, design in "aided" systems, or automate the function). Any attempt at creating a biocybernetic system will need to take into account the demand characteristics of the workload prior to being able to feedback such information so that the system produces less workload.

We are aware that there is a need to do better at workload measurement, but we would suggest that the improvement should be not only from the standpoint of the workstation, but also how it affects the individual operator. This is not to say that we believe that subjective measures of workload are necessary but, rather, we emphasize that those which are subject (or operator) dependent should be studied more extensively. Our bias is that the physical characteristics of the stimulus are likely to be predictive of the general level of workload, and so subject independent methods of workload assessment are necessary. However, such indices are chiefly useful to create specifications. Because so much of the reliable variance in operational performance is identifiable as being due to the individual (Kennedy, Jones, & Baltzley, 1988), there is also a need for subject dependent measures. Note that both of these (subject dependent and subject independent) can be objective measures (e.g., physiological indicants) or subjective (e.g., self-report). We believe that the present study will deal with objective measures which can be employed for both subject dependent and subject independent assessment. This classification schema and the implications for workload measurement are described in greater detail elsewhere (Kennedy, May, Jones, & Fowlkes, 1989).

Whatever method is followed, workload metrics are needed that can be fed back to operators to remediate performance decrements. This paper takes the position that an individual's difficulty with a work station -- whether due to time-on-task, overload or some other factor -- is reflected, perhaps unconsciously by some neuroelectric events which are recoverable and measurable. It is our intention to capture these events and determine ways in which they can be assessed portably and quickly so that their predictions can be fed back to the system with which the individual interacts in order that the system will better accommodate the individual's needs. Following Wiener (1948), we think that something like this is what K.U. Smith (1966, 1967; Smith & Smith, 1966) had in mind when he began writing about biocybernetics some 25 years ago.

RELATED RESEARCH

Impetus for the present effort began several years ago with the experimental observation that performance on a vigilance task was correlated with patterns of eye movements. Specifically, the fast phase of vestibular-induced nystagmus appeared to disappear as performance (percent correct) degraded over a sustained period of time (Kennedy, 1972). This might be expected, it was argued, since the fast-phase component of nystagmus is dependent on the integrity of the reticular formation which is also related to arousal and alertness (Cohen, Feldman, & Diamond, 1969; Darhoff & Hoyt, 1971; Yules, Krebs, & Gault, 1966). Given that other research related nystagmus to arousal (Collins, Crampton, & Posner, 1961; Collins & Posner, 1963), the idea was also pursued in a Navy project for divers (Kennedy, 1978) where velocity was shown to be related to both performance and time on task.

More recently, under sponsorship of AFOSR, experimentation of an eye movement metric of workload and arousal was found where in two studies (May, Kennedy, Williams, Dunlap, & Brannan, 1985) subjects performed a complex tone counting task at three levels of difficulty while the velocity and latency of saccades in the horizontal plane were recorded via infrared reflection techniques. In that study saccade latency appeared to vary with the task demands but practice effects and individual differences in performance masked the relationship. There was a suggestion that the length of a saccade may provide information about task demands. Support for the "velocity hypothesis" was not found -- saccade velocities obtained during the baseline condition of free-viewing did not differ significantly from velocities obtained under tone counting conditions -- however, there were known limitations in sampling rate and recording sensitivity of the technique employed.

In the second project (Kennedy, May, Jones, & Fowlkes, 1989) the same equipment and task were used to further examine the relationship of eye movement measures (saccade length) to workload. The general finding was that the mean range of eye movements in the horizontal plane for each of the three workload conditions varied inversely with difficulty level of the tone counting task. Thus, the results from this research suggested that saccade length could be a promising objective measure of the task demands of a display and thereby serve as a useful measure of mental workload.

DESCRIPTION OF PROJECT GOALS

The purpose of the present study was to investigate eye movement metrics which would covary with task loading in order to derive an empirical index of workload. Additionally, because such metrics are likely to reflect the task demands, it was predicted that changes in the metrics, which occur over sustained periods of time, could serve to signal loss of attention or vigilance. An ancillary technical obstacle to solution of such problems is the requirement for specially customized scoring and analytic techniques. We therefore set out to address three goals collectively which are summarized below. The results for the first two goals are reported in the experimental section and progress made in the development of automated scoring techniques is presented in the section on software development.

1 Compare eye movement metrics to performances on psychophysically scaled workload tasks. Establish the relationship between the eye movement metrics and taskload and task performance.

2 Examine the effects of sustained time on task on the eye movement metrics.

3 Develop computerized techniques and algorithms (hardware and software included) for rapid scoring and analysis of eye movement indicants of workload and vigilance.

The ultimate outcome of Phase II would be the prototype of a bundled transportable system to assess mental workload via eye movement and other bioelectric measures as appropriate. A chief ingredient in initial development of such a system would be the focus on rapid (i.e., seconds) evaluation of the bioelectric events so that, when properly identified and classified, such signals could be fed back to signal generators.

III. EXPERIMENTAL EFFORT

METHOD

SUBJECTS

College students were recruited from the University of Central Florida and were paid $5 per hour for their participation. The subjects received informed consent forms and were otherwise used in accordance with established policies of human use according to nationally published guidelines. All subjects were instructed to perform low, medium, and high difficulty levels of an auditory tone counting while eye movements were recorded using electro-oculography (EOG).

EQUIPMENT

Four mm silver/silver chloride electrodes were used for EOG recording. Electrode leads were fed to three amplifiers featuring characteristics suitable for EOG recording. A MetroByte Corporation DAS-16F 8 channel (bipolar) analog-to-digital converter capable of sampling up to 100,000 samples per second to serve as the interface to the microprocessor. Three channels were sampled, two for vertical eye movements and one for horizontal eye movements. Each of the three channels were sampled at 256 samples per second each (i.e., 256 samples per second (SPS) for the horizontal channel, 256 SPS for the left eye vertical movement, 256 SPS for the right eye vertical movement).

The calibration board contains 8 red LEDs imbedded in a four foot square panel which was painted black. The LEDs are controlled via software using the 8 digital I/O channels provided by the DAS-16F. With the subject seated 5 feet from the center of the board, and at eye level, 40 degrees of vertical and 40 degrees of horizontal distance separate the top/bottom, left/right LEDs, with 10 degrees of separation between each LED. The software calibration routine successively illuminates the calibration LEDs in the following order: -20 degree vertical, -10 degrees vertical, +10 degrees vertical, +20 degrees vertical, -20 degrees horizontal, -10 degrees horizontal, +10 degrees horizontal, +20 degrees horizontal and 0 degrees (center LED). During calibration and while performing the tone counting task, head movement was restrained using a chin block.

PROCEDURE

Upon arrival at the experimental site, all subjects were briefed regarding the purpose and procedures of the experiment and then were given informed consent forms to read and sign. Three separate efforts were conducted. In the first several subjects were exposed to the experimental procedures in order to facilitate technique and to evaluate equipments, scoring, etc. The main experiment employed 20 subjects but various difficulties (power outage, scheduling, etc.) resulted in complete data to be available for 15. The third study, a probe, employed three subjects over a one hour period of recording/testing.

Vertical and horizontal eye movement recordings were taken for all subjects. Following the initial briefing, skin areas where the surface electrodes were to be applied were cleansed and then electrodes were attached (see Figure 1) and impedances checked.

The complex visual counting task, developed by Jerison (1956), was employed to manipulate workload. This task has since been modified for auditory presentation (Kennedy & Bittner, 1980) and has been used for many years (Kennedy, 1971). In the first study, subjects (N=15) were instructed to perform low, medium, and high difficulty levels of the auditory tone counting while eye movements were recorded. In the third study, only the two channel task was employed. A microcomputer was used to administer a quasi-random series of high, middle, and low pitch tones. The subject's task was to retain the count of tones in working memory. Under the low workload (one tone) task the subject counted each occurrence of the low pitch tone and depressed a key after each fourth occurrence. The medium workload task required that the subject hold in working memory both the low and middle pitch tones separately even though they were presented at different rates (two tones). Low, middle, and high pitch tones were counted and held in memory separately for the high workload condition. Because initial pilot testing indicated that subjects (N=15) were making minimal eye movements during tone counting, it was decided to add the a visual component to the task so that the occurrence of the left, center, and right LEDs (separated by 10 degrees horizontally) corresponded to the low, middle, and high pitch tones. Therefore, subjects were told to monitor the lights in addition to the tones.

A schematic arrangement of the test procedure for task practice, calibration, presentation of the workload conditions is presented in Figure 2. Prior to the start of data collection, subjects received 2.5 minutes of practice on each task. Following practice, each of the three workload performance tasks was administered twice for 2.5 minutes while eye movements were measured. To counterbalance order of administration, approximately 1/3 of the subjects received each of the three workload conditions first, second or third. The total session length was approximately 40 minutes.

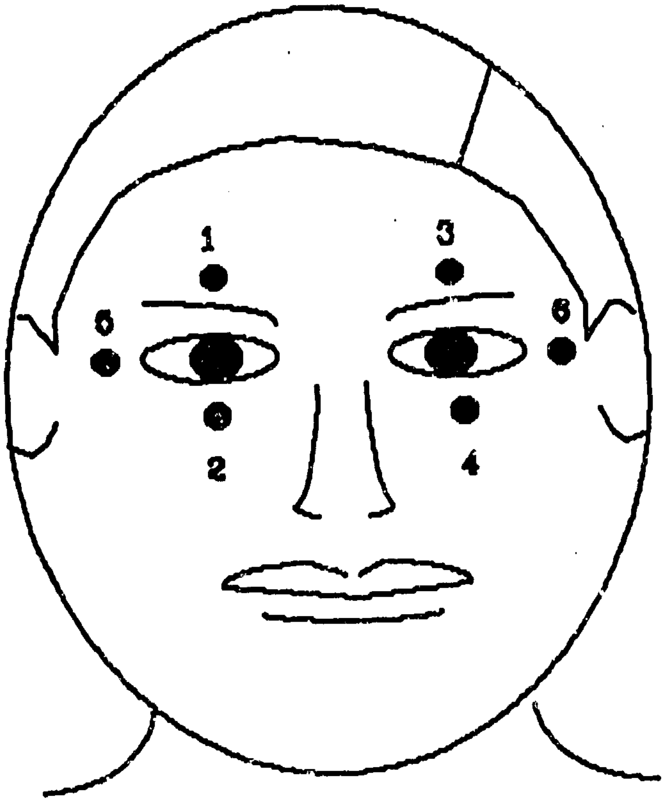


Figure 1. Placement of Surface Electrodes for Eye Movement Recordings (ground electrode attached to inner left wrist not shown)

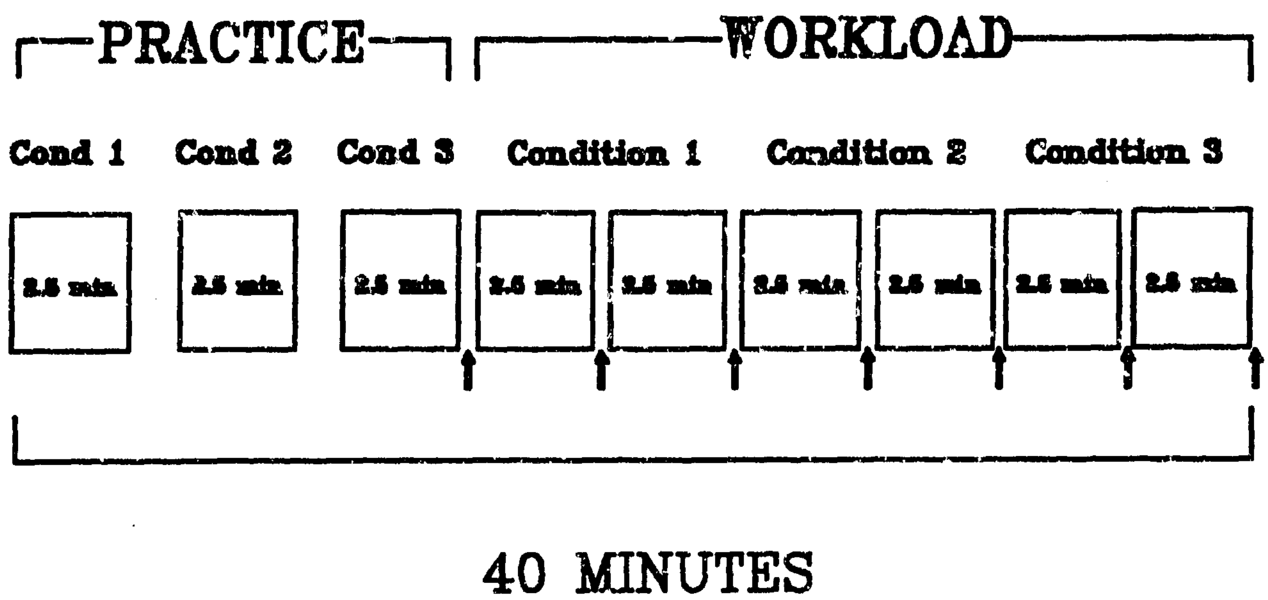


Figure 2. Schematic Arrangement of the Test Procedure for Practice, Calibration and Test Administration.
(" " indicates calibration)

RESULTS AND DISCUSSION

PERFORMANCE DATA - MAIN STUDY

Figure 3 shows performance averaged across 15 subjects for each of the three levels of workload. As predicted, performance (percent correct) was best on the "low" workload task and poorest for the "high" workload task and these differences were statistically significant ($p < .01$). On the high workload task, no subject obtained a perfect (100% correct) score for the five minutes of performance. On the low workload task over half of the subjects obtained 100% correct for the five minutes of testing. This implies that this procedure was successful at the group level in producing task control of workload.

The individual summary data for each subject are shown in Table 1. In general the individual subject's performances reflect the group relations although individual differences in responding by some subjects implies they may have used different strategies in their approach to the "high" versus "low" workloads of the experimental situation. This means that while group relations may provide statistically meaningful relationships, in subject dependent performances such as these, assessment of individual bioelectric need to be studied as well.

Although session length was only 40 minutes and attempts were made to maintain motivation for the task, some of the subjects observed that the tone counting tasks, particularly for the high workload condition, were considered taxing. Therefore we compared early versus late performances for evidence of a vigilance decrement. Predictably, tone counting performances were generally better early in the session, than later, but these differences were not significant (cf., Figure 4). However, because of the

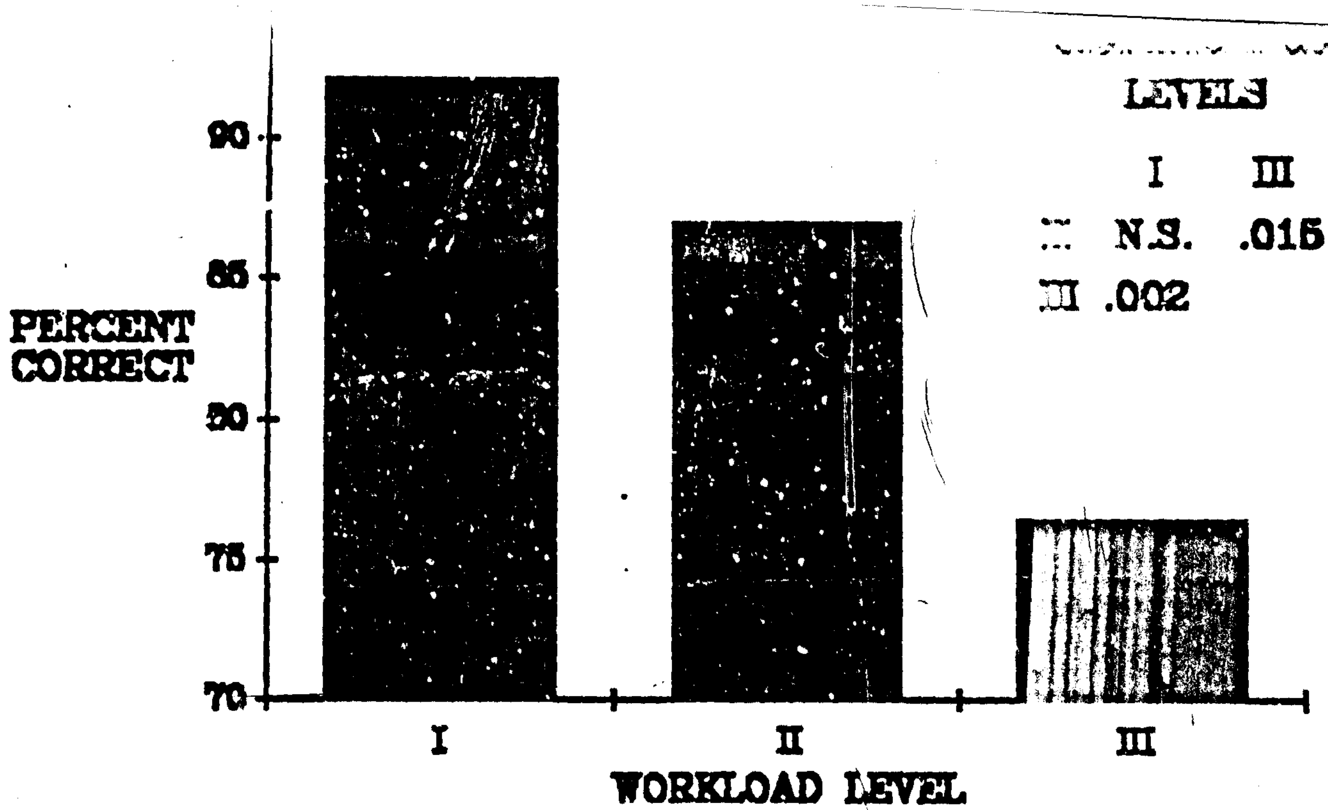


Figure 3. Three Levels of Workload for 15 Subjects

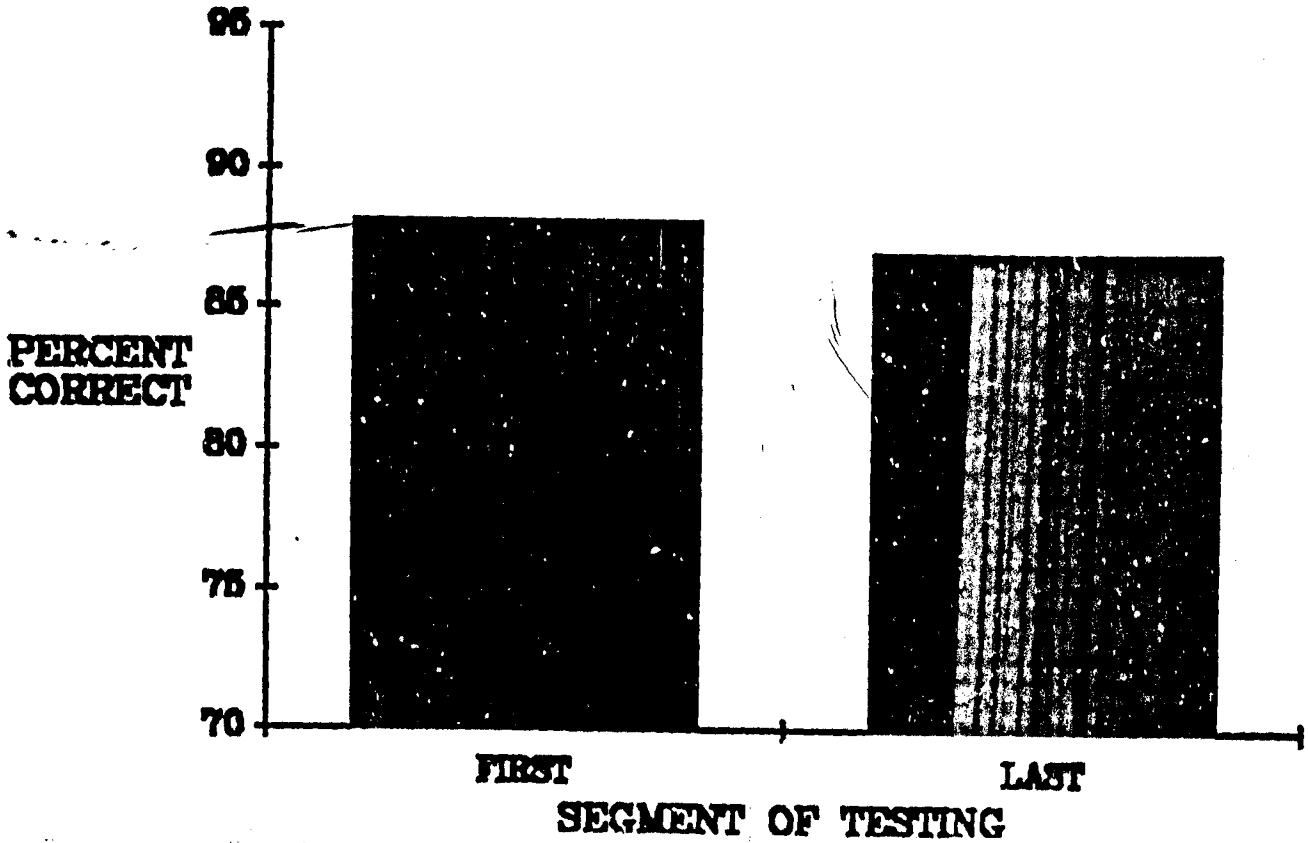


Figure 4. Performance Scores for the First and Last Segment of Testing

TABLE 1. Percent Correct Response for Each Subject for the Low, Medium, and High Workload Conditions Along with Summary Statistics

Sub	Low	Med	High	Sub	Low	Med	High
xxh	92	91	92	dsk	75	100	63
bms	100	77	50	ehw	92	40	67
chb	100	91	77	bam	75	95	77
jmc	100	91	77	cgg	100	91	90
das	100	100	67	wow	83	86	73
jat	100	86	90	dat	75	100	87
bsw	92	100	87	wjt	100	91	70
teh	100	68	60				

SUMMARY STATISTICS

	Workload Low	Med	High
Mean	92.27	87.13	75.13
(SD)	(10.21)	(15.80)	(12.51)

PAIRED T-TESTS

LOW VERSUS MEDIUM: $t(14)=0.95$, $p=.357$
LOW VERSUS HIGH: $t(14)=3.93$, $p=.002$
MEDIUM VERSUS HIGH: $t(14)=2.94$, $p=.015$

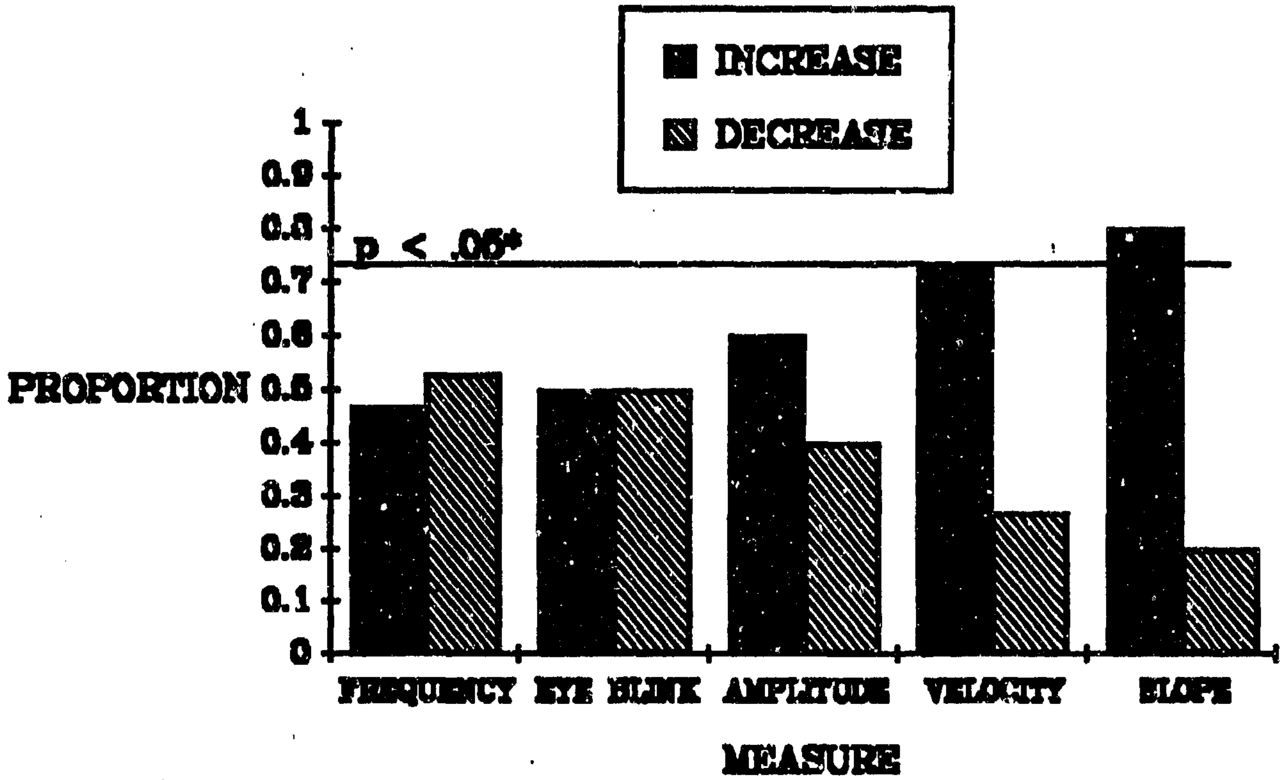
experimental set up and the requirement for counterbalancing, minimal power could be expected for testing this relationship, and this may have masked statistical verification of such expected effects. Elsewhere for sustained performance measures, longer sessions have been employed (Kennedy, 1972) and in those cases operators exhibit worse performance later in the session than earlier. In the longer sessions of the study reported below we found the expected performance decrement with time on task.

DEPENDENT VARIABLES - EYE MOVEMENTS

To facilitate comparisons in evaluating the various eye movement indicants, we have treated the one and three tone monitoring tasks as "low" and "high" workload conditions and have summarized the findings in Figure 5, which shows the proportion of subjects who showed an increase or decrease in each dependent measure from the low to the high workload condition. Additionally, to index possible time course eye movement metrics, we have made summarized comparisons for the first five minutes of exposure versus the last five minutes in Figure 6 (showing the proportion of subjects who showed an increase or decrease in each dependent measure from the first to last five minute segment of testing). This approach to the examination of the workload metrics combines data over the two 2.5 minute sessions for each condition. In the findings reported, the same general patterns were generally obtained in both 2.5 minute sessions, and the sessions, therefore, constituted a replication. In future work, subjects should be tested over replications conducted on separate days as a check on the reliability of the findings.

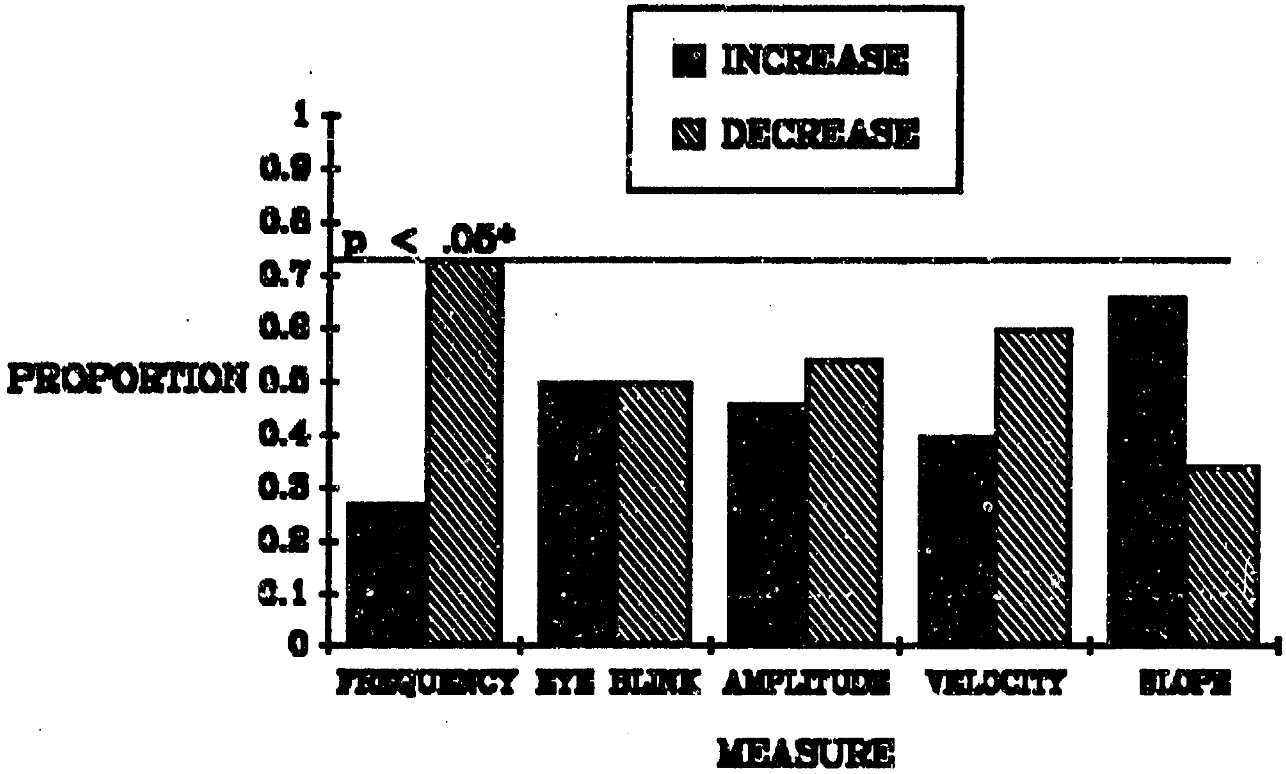
EYE BLINKS

Eye blinks have been shown to be related to workload. Stern and Skelly (1984) assessed eye blink, blink rate, and blink duration in two simulated bomber missions. Differing task demands were reflected by blink rate and, in addition, blink duration appeared to be affected by time on task. In other research, Bauer, Goldstein, & Stern (1987) found that, in an encoding and retention task, blink rate was depressed following the presentation of task cues, memory set, and test stimulus. The findings for blink rate from the present research are summarized in Figures 5 and 6 and presented for individual subjects in Tables 2 and 3. In this study, although blink rates appeared to be reliable within a subject, blink rate *per se* does not appear to be affected by task demands nor was blink rate affected by time on task.



*Based on test (1-tail) for significance of a proportion.

Figure 5. Proportion of Subjects Showing an Increase or Decrease in Each Measure From Low to High Workload.



*Based on test (1-tail) for significance of a proportion.

Figure 6. Proportion of Subjects Showing an Increase or Decrease in Each Measure From the First to Last Segment of Testing.

TABLE 2. Number of Eye Blinks for Each Subject for the High and Low Workload Conditions

Sub	Low WL	High WL	Sub	Low WL	High WL
xxb	52	22	dsk	64	69
bms	8	14	ehw	11	4
chb	24	13	bam	22	19
jmc	63	155	cgg	16	19
das	41	23	wow	42	43
jat	13	8	dat	*	*
bsw	138	179	wjt	12	17
teh	10	9			

* No eye blink data collected.

TABLE 3. Number of Eye Blinks for Each Subject for the First and Last Five Minutes of Testing

Sub	1st 5 Min	Last 5 Min	Sub	1st 5 Min	Last 5 Min
xxb	31	52	dsk	69	73
bms	14	12	ehw	11	4
chb	29	24	bam	21	22
jmc	63	68	cgg	36	19
das	23	41	wow	91	43
jat	13	20	dat	*	*
bsw	138	179	wjt	22	17
teh	10	9			

* No eye blink data collected

FLICK FREQUENCY

There were wide individual differences in the number of saccades made by a subject per five minute epoch of workload performance and these differences were very reliable. Frequency of saccades appear in Table 4 for the two workload conditions. It may be seen that they do not vary systematically as a function of workload (see Figure 5). However, Table 5 which contains frequency of saccades for all subjects for the first and last five minutes of testing shows that for 11 out of the 15 subjects (73%; $p < .03$), there was a decrease in frequency of saccades from the first to the last five minutes of testing (see Figure 6). Frequency of saccades therefore may be related to time on task, and should be followed as a prospective measure of the vigilance decrement and sustained performance.

TABLE 4. Number of Saccades for the Low and High Workload Conditions for Each Subject

Sub	Low WL	High WL	Sub	Low WL	High WL
xxb	267	195	dsk	153	124
bms	41	73	ehw	168	152
chb	311	306	bam	93	338
jmc	198	222	cgg	23	18
das	131	178	wow	24	11
jat	15	16	dat	240	230
bsw	221	84	wjt	17	24
teh	64	92			

TABLE 5. Number of Saccades for the First and Last Five Minutes of Testing for Each Subject

Sub	1st 5 Min	Last 5 Min	Sub	1st 5 Min	Last 5 Min
xxb	338	267	dsk	124	134
bms	73	29	ehw	168	152
chb	212	311	bam	393	93
jmc	189	171	cgg	41	18
das	178	131	wow	47	11
jat	15	16	dat	240	230
bsw	221	84	wjt	17	24
teh	92	34			

AMPLITUDE OF SACCADES

In previous work, under AFOSR sponsorship we (Kennedy, May, Jones, & Fowlkes, 1989) found the size (i.e., length) of a saccade was related to workload, such that the mean range of eye movements in the horizontal plane for each of the three workload conditions varied inversely with difficulty level of a tone counting task. The present study did not find support for this finding. It may be seen in Figure 5 average amplitude of saccades was shown to be greater for the high versus the low workload conditions (data for individual subjects are presented in Table 6). It should be pointed out that the conditions of the present experiment were slightly different from those employed previously and entailed somewhat greater visual articulation for the high as opposed to low workload conditions. That is, there may have been greater requirements for visual activity in order to monitor the LEDs which were coincident with the auditory tones and this may have influenced eye movement activity. Thus, under the difficult tone counting condition, subjects may have been more likely to monitor all three lights whereas in the low workload condition subjects may have been more likely to monitor only one LED (corresponding to the occurrence of the low pitch tone). Additionally, since performance data in this experiment (percent correct) were higher than in our previous research (possibly due to the use of redundant information - lights + tones - in this study), an alternative interpretation is that the task demands in the high workload condition were not strong enough stimulus to induce a restriction in saccades. These findings underscore the requirement to specify very completely the content of the visual scene when eye movement related metrics are studied since they are likely to interact.

Figure 6 (data for individual subjects presented in Table 7) summarizes the fact that there was little change ($P < .15$) in amplitude as a function of time on task; However, the small apparent trend there was in support of the SLIT prediction (i.e., a tendency toward larger saccades over time on task).

TABLE 6. Average Amplitude of Saccades (in degrees) for Each Subject for the High and Low Workload Conditions

Sub	Low WL	High WL	Sub	Low WL	High WL
xxb	5.39	5.33	dsk	9.21	8.08
bms	5.70	5.70	ehw	9.70	11.51
chb	3.75	7.02	bam	6.26	7.33
jmc	15.28	10.49	cgg	6.93	9.72
das	8.60	8.68	wow	2.39	8.37
jat	3.98	4.14	dat	13.63	13.93
bsw	8.17	6.97	wjt	8.05	8.21
teh	4.08	3.34			

VELOCITY OF SACCADES

Average velocity of saccades shown as a function of workload and time on task are shown in Tables 8 and 9 and are also summarized in Figures 5 and 6. From these summary figures it may be seen that the original hypotheses receive support. That is, velocity of saccades generally showed an increase from low to high workload and a decrease from the first to last five minute segment of testing. It is also apparent from the figures that the trends for velocity paralleled the trends seen for amplitude. Because larger amplitude saccades are associated with greater velocities, we believe that the trends for

TABLE 7. Average Amplitude of Saccades (in degrees) for Each Subject for the First and Last Five Minutes of Testing

Sub	1st 5 Min	Last 5 Min	Sub	1st 5 Min	Last 5 Min
xxb	6.82	5.39	dsk	8.08	9.11
bms	5.70	5.23	ehw	9.70	11.51
chb	5.64	3.75	bam	6.00	6.26
jmc	15.28	11.38	cgg	10.34	9.72
das	8.68	8.60	wow	6.85	8.37
jat	3.98	3.03	dat	13.63	13.93
bsw	8.17	6.97	wjt	5.65	8.99
teh	3.33	4.08			

velocity could be due to the constraints of the task - subjects would be more likely to make larger amplitude saccades, and therefore saccades of greater velocity, in the high workload condition.

TABLE 8. Average Velocity of Saccades (in degrees/second) for Each Subject for the High and Low Workload Conditions

Sub	Low WL	High WL	Sub	Low WL	High WL
xxb	145.26	154.08	dsk	220.29	207.72
bms	179.73	198.28	ehw	237.63	280.39
chb	104.61	177.58	bam	148.62	185.75
jmc	378.91	276.91	cgg	182.08	226.59
das	160.47	181.33	wow	73.92	218.44
jat	74.56	101.53	dat	262.47	267.81
bsw	194.19	175.89	wjt	215.30	255.19
teh	108.97	92.55			

TABLE 9. Average Velocity of Saccades (in degrees/second) for Each Subject for the First and Last Five Minutes of Testing

Sub	1st 5 Min	Last 5 Min	Sub	1st 5 Min	Last 5 Min
xxb	187.80	145.26	dsk	207.72	221.18
bms	198.28	164.98	ehw	237.63	280.39
chb	145.74	104.61	bam	160.85	148.62
jmc	378.91	274.69	cgg	232.63	226.59
das	181.33	160.47	wow	193.71	218.44
jat	74.56	65.93	dat	262.45	267.80
bsw	194.19	175.89	wjt	102.63	255.19
teh	92.55	108.97			

A NEW ACCELERATION METRIC

Because of the high degree of association between amplitude and velocity (e.g., $r > .85$ in our sample), it was deemed advisable to control further for this relationship when examining velocity. Simply stated, because the orbit of the eye limits the extent over which eye movements may be made, and because all eye movements follow the general form of an acceleration followed by a deceleration, it was felt that if there were a triggering command signal, as was hypothesized, it would first be necessary to control for the physical mechanics of the eye movement itself. Therefore, to further analyze this hypothesis, the method of least squares was used to fit a linear regression line to the function between amplitude and velocity so that the steepness of the slopes for the high and low workload conditions could be compared. Logically then, the difference between the steepness of those two slopes would signify relatively greater rates of velocity (i.e., acceleration) between the two conditions. A scatterplot of velocity with amplitude is shown for two subjects in Figure 7 with the best fit linear regression lines for the high and low workload conditions also shown. It can be seen that a steeper slope was obtained in the high workload condition for both subjects. This finding was obtained for 12 of the 15 subjects (80%), as shown in Figure 5, suggesting an increase in rate of velocity (i.e., acceleration) in the high workload condition.

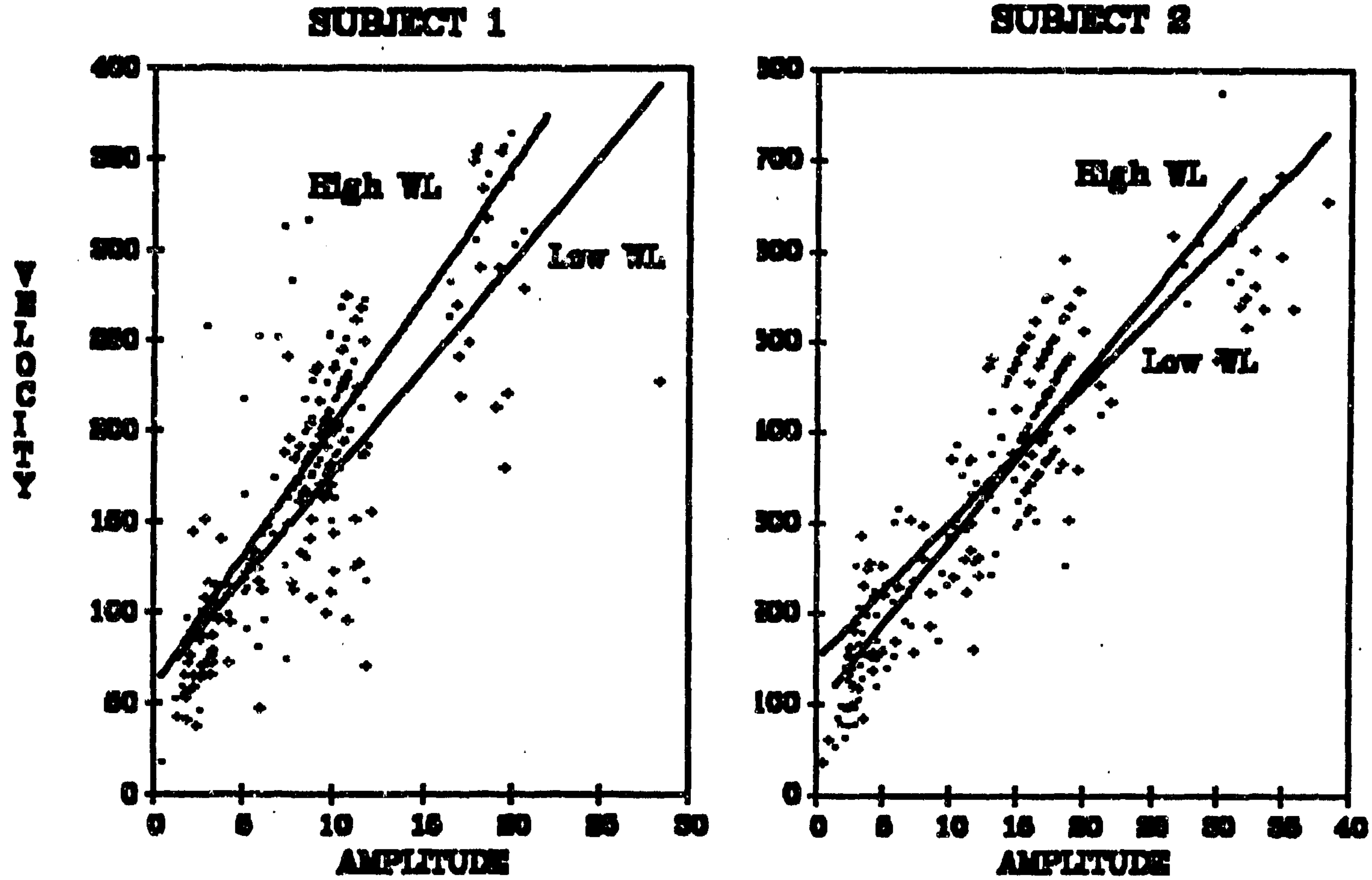


Figure 7. Linear Regression Lines for Low and High Workload Conditions for Two Subjects.

PILOT STUDY IN SUSTAINED PERFORMANCE

In a pilot study, using only the middle workload condition for the tone counting task, two subjects were tested continuously for 21 minutes while eye movements were recorded without benefit of calibration. The purpose of this probe was to determine whether a longer period under one task loading and uninterrupted by calibrations would result in a vigilance decrement and thereby serve as a model for sustained performance studies. The secondary purpose was to compare the eye movement metrics over a protracted period of time with only one level of workload. The primary purpose was largely successful. Data recording and analysis of eye movement was accomplished without incident. The performance scores were less clearcut. For one of the subjects, performance for the first seven minutes was perfect (100% correct), while it decreased to 73% and 80% correct for the second and third seven minute epochs, respectively. Also, from the eye movement analysis, a similar relationship between apparent workload (indicated by performance) and change in slope was observed as was obtained in the workload study. Over this period of sustained performance, there was a systematic increase in slope scores from the first to the last period, suggesting that, as alertness decreased through time on task, there may also be a consequent increase in the mental demands of the task due to the tedious nature of the monitoring activity. Stated differently, to the extent that these slope scores index task demands, time on task may have an effect similar to adding additional channels of information to be monitored. There was also a systematic decrease in the number of saccades as a function of time on task, as was found in the workload experiment. The data from the second subject, while satisfactorily obtained, are more difficult to reconcile. For example, performance scores were poor throughout the period (67%, 60%, and 70% correct for the first, second, and third periods) and showed negligible change over session. Although frequency of saccades for the second subject decreased as a function of time on task, slope scores systematically decreased. Therefore with regards the slope metric and time on task, the findings from this probe remain inconclusive.

IV. SOFTWARE DEVELOPMENT

As part of the present effort we sought to develop computer techniques and algorithms for rapid scoring and analysis of eye movement data. In previous work for the Air Force, we developed considerable amounts of customized software which will permit automatic scoring of the eye movement data. The automatic scoring program correctly identified approximately 60% of saccades and eye blinks. Under the present contract, we have further refined this software so that it is now interactive. This was seen as a necessary intermediate step in the development of a fully automated scoring system.

DESCRIPTION

We utilize FORTRAN-IV object modules to aid in the sampling of Analog-to-Digital channels, the programming of the on-board Real-Time-Clock, and programming of the Digital I/O channels. This library is used extensively in each of the software programs described below.

We have written software to permit storage of analog-to-digital conversions of eye movement signals (including two-dimensional recording), data archiving, and immediate play-back. In addition, the software takes a novice system user through calibration techniques and actual use of the system.

While it would have been desirable to have immediate on-line scoring of saccade lengths, eye blinks, and saccade durations, it was found that with the computations involved, (even with the high-speed processor in use), it was not feasible without an increased expenditure for more expensive and faster hardware.

Currently, we are able to analyze a two and a half minute session and have the results stored in an intermediate file in less than two minutes. With the price of fast DRAM chips coming down to a reasonable figure, and with increased availability, most 80386 machines are now within the price range of our system; and we would anticipate that concurrent processing of the data collected during a session is now a fiscally prudent possibility.

DESCRIPTION OF THE SOFTWARE MODULES

All of the executable modules are kept on a single partitioned drive of the computer. The source code for the Essex developed modules are located in a sub-directory as are the libraries used for linkage. Data are collected in the main directory, and after an experimental run, are 'squashed', archived, and stored in a DAT sub-directory.

There are four modules which need be executed to cycle through an experimental run: LIVESHCW, BIOC, SQUASH and SCORE.

LIVESHOW is run prior to the start of data recording, and after the electrodes have been tested. Because there tends to be a slight drift in the signals, from the beginning of a session to the end, it is necessary to ensure that all signals start on the zero line. LIVESHOW enables to experimenter to center each signal prior to the execution of the experiment by displaying the incoming signals from the subject on the EGA monitor.

Once the hardware has been calibrated, the run-time data recording task, BIOC, is executed. Upon execution, the experimenter is prompted for a three letter subject identifier, most often the first letter of the subjects first, middle and last name. The experimenter then enters the order of the tasks to be performed by the subject, (i.e., high, middle, or low demand).

Each workload level is broken down into two 2.5 minute sessions. A software calibration period precedes and follows each 2.5 minute session. It was decided to place a calibration section at the end of a session due to the slight drift that occasionally occurs. With a calibration period at the beginning and the end of each session, automatic compensation can take place in the scoring module. The software calibration routine successively illuminates the calibration LEDs in the following order: -20 degree vertical, -10 degrees vertical, +10 degrees vertical, +20 degrees vertical, -20 degrees horizontal, -10 degrees horizontal, +10 degrees horizontal, +20 degrees horizontal and 0 degrees (center LED). The subject is instructed to fixate on the illuminated LED and one second's worth of data is collected and stored in the data file.

The next 150 seconds worth of data consist of the actual eye movement data, followed by 9 seconds of calibration data and, finally, the performance scores for the trial session. Data are stored as 2's complement two-byte integer values, using the upper 12 bits of the word. The first word of the file is data for A/D channel 0, the second is channel 1 and the third word is channel 2. Data are stored in the above order for each sample taken during the run. Since we sampled at 256Hz, one second worth of data takes 1536 bytes.

After the last workload level is completed, (when the experimental run is over), the SQUASH module is executed to compress the file. Due to the timing of writing to the disk, sometimes zero data is transferred into the file. SQUASH removes this non-data and effectively cuts file size by approximately 23 percent. Since a complete testing session generates over 2.5 megabytes of data, SQUASH effectively reduces this amount to nearly 2 megabytes. After scoring the data, we further compressed the data using PKARC, an archiving program, which reduced the size of the data files to about 1.25 megabytes, and then we moved the archive to the DAT subdirectory.

The SCORE module prompts the experimenter for the three letter subject ID, the type of graphics monitor connected to the computer (e.g., CGA, EGA, VGA), the sample rate used during data collection, and the number of data points to be used for tentative saccade initiation. After reading in the initial nine seconds of calibration data, the program becomes interactive. During this experiment, the data was collected using a sample rate of 256Hz. On most PC's, this allows the display of two seconds worth of data on one screen. The actual data collected is then displayed on the monitor, scored, and results stored in a filename composed of the first three initials of the subject's name, the testing session, and the extension 'SAC'. The SCORE program displays the two vertical channels in dark blue and light blue, while the horizontal channel is graphed in green. For each frame of data displayed (2 seconds), SCORE marks any blinks or saccades which the program has attempted to identify. Blinks are marked with a vertical light blue line, and saccade beginnings and endings are designated by green vertical lines at the beginning and end of the saccades identified by the program. At this point, the experimenter has the option of accepting the opinion of the program, or, rejecting the computer generated detections by either adding or removing saccade and blink identifications. By using the arrow keys on the keyboard, the experimenter can move a vertical red line serving as a cursor and by positioning this cursor where desired, either add or remove any blink or saccade detected. Table xx shows the functions of the keypad. Results of the scoring session are stored in the 'SAC' file in the following format:

event type, event datum1, event datum2, event datum3, event datum4

where event type would denote:

0 - Initial calibration data
1 - Computer detected blink
2 - User detected blink
3 - Computer detected saccade
4 - User detected saccade
5 - Ending calibration data

Depending upon the type of event, the data following consists of varying values. For blinks (computer or human detected), the first value contains the data point number of the middle of the blink; the other three values are null values (i.e., they don't carry any meaning). All of the four values for saccades are meaningful, on the other hand. The first value contains the data point number of the initiation of the saccade, the second the data point number of the termination of the saccade, and the third and fourth contain the voltages of the beginning and end of the saccade, respectively.

After the experimenter has scored the subject's data, another program is run which computes and stores summary data for each session. This file is named using the subject's three initials and a 'SCO' extension. Reported for each session are number of eyeblinks, number of saccades, average duration, and the average velocity (degrees per second) of each saccade. The same information is reported for left and right saccades. The final listing displays subject performance on the counting task.

TABLE 10. Description of Scoring Keys.

Left Arrow	- Move cursor left one unit
Shift Left Arrow	- Move cursor left 10 units
Control Left Arrow	- Move cursor left 50 units
Right Arrow	- Move cursor right one unit
Shift Right Arrow	- Move cursor right 10 units
Control Right Arrow	- Move cursor right 50 units
Home	- Position cursor at first position
End	- Position cursor at last position
Up arrow	- Move horizontal display up 10 units
Down arrow	- Move horizontal display down 10 units
S or s	- Beginning of saccade command sequence
S or s	- Mark beginning of saccade
E or e	- Mark end of saccade
R or r	- Remove marked saccade
B or b	- Beginning of blink command sequence
S or s	- Mark blink at current cursor position
R or r	- Remove blink at current cursor position
Escape	- Terminate current command sequence
?	- Expand channel displays
!	- Contract channel displays
Enter	- Accept and store current screen

Computer saccade identification is comprised of three stages: the first identifies saccade initiation, the second saccade termination, and the third computes the voltage delimited by these points to arrive at the size of the saccade. To identify saccade initiation, the program calculates the absolute difference in amplitude between five successive samples (19.5 ms) from the horizontal channel. If all five differences exceed 3.66 millivolts AND are in the same direction, a tentative saccade initiation is flagged. Saccade termination is determined by searching for five successive samples whose differences are less than 3.66 millivolts. If saccade termination criteria is not met within 200 ms, the tentative saccade is flagged as a false saccade, and not included in the data analysis.

The size of the saccade is computed by taking the absolute difference in amplitude of the end of the saccade from the beginning of the saccade. This results, in millivolts, is then multiplied by the scale factor obtained from the calibration routines to arrive at the number of degrees the eye moved. Duration of the saccade is computed by simply subtracting the number of the data point which began the saccade, from the ending data point number and multiplying by the known sample rate, which in this case, is 256Hz. For example, if a saccade began at data point number 1432, and ended at data point number 1466, the duration of the saccade would be 34 x .0039063, giving a result of 132.8 milliseconds. Velocity of the saccade, in degrees per second, is estimated from dividing the score obtained above by the duration of the saccade.

Eye blinks are relatively easier to detect. The vertical channels are monitored for quick, positive successive slopes. If five successive samples show this trend, a tentative blink flag is set. The termination of a blink is detected by checking for five successive samples with quick, negative slopes.

V. CONCLUSIONS

The idea which prompted the present research was that biological events may be predictive of the attentional and task demands of work. If these could be analyzed in real time and fed back to the machine (or operator), a truly biocybernetic system could be created. The purpose of Phase I was to determine the feasibility of using eye movement data to provide an indication of operator mental state.

In Phase I we set out to accomplish three goals:

1 Compare eye movement metrics to performances on psychophysically scaled workload tasks. Establish the relationship between the eye movement metrics and taskload and task performance.

2 Examine the effects of sustained time on task on the eye movement metrics.

3 Develop computerized techniques and algorithms (hardware and software included) for rapid scoring and analysis of eye movement indicants of workload and vigilance.

For the experimental effort (goals 1 and 2 above), we demonstrated, first, that task control of workload had been obtained. Thus, percent correct scores obtained in this study covaried with objective (number of tones) indices of workload. Further, eye movement metrics bear a relationship to these scaled values so that amplitude, velocity and more strongly, acceleration of eye movements appear to be related to these objective controls of workload. Moreover, 80% of the subjects show the relationship between the acceleration metric and workload. The possibility exists that if personal performances were assessed, it may be possible to tune the acceleration metric in order to improve on its predictive power.

The case for time on task is somewhat weaker because the period of sustained performance was short and the change in performance due to a vigilance decrement was small. However, again the acceleration score shows a change (it gets faster with time on task) which implies that it is measuring some combination of task loading plus time on task. The frequency of eye movements also decrease with time on task. Both metrics should be followed as a prospective measures of the vigilance decrement and sustained performance. A pilot study with only two subjects and a single workload task demand produced mixed results: parallel findings for one subject who showed the expected decrement in task performance as a function of time on task and contrary findings in a subject whose performance was low and did not degrade. These relations require further study with more subjects, over longer periods of performance and who can receive disparate task controls and demands and mental work.

In summary, neither two, nor even fifteen subjects may be considered sufficient to do more than gather hypotheses for further study. To some extent the findings reported above are encouraging. Both eye movement velocity and velocity controlled for amplitude (what we called acceleration) appear to be related to the objective task demands on a subject. This is now an empirical outcome in this study and follows from theory presented elsewhere and reported above. Thus when more stimuli are presented to be retained in working memory, this produces a change in eye movement character so that the acceleration (or velocity) is relatively greater in a subject under the high, rather than the low, workload condition. We hasten to point out that this relationship is a relative one and requires that an individual first be calibrated for his/her own "signature" (obtained from a lower level of workload) in order to compare to later outcomes with increased task demands. We believe that there is considerable strength in this differential approach to biomedical research. Specifically it suggests that power (statistical and otherwise) can be garnered from using each subject as his/her own control and then later on customize procedures based on this individual knowledge. In the present study, using group specific findings (viz., means and standard deviations), none of the outcomes would have been revealed.

Beyond the statistically significant outcomes there are other relationships which pose more questions than they answer. For example, does the obtained acceleration metric assess mental workload, arousal, attention, working memory, lack of boredom, etc.? Why does the acceleration metric correlate positively with velocity during workload and negatively during time on task? Is it possible (or likely) that time on task is likely to produce a reduction in attention in some subjects and do any of these metric reflect that relationship? Is frequency a function of time on task or is it related to mental/neural activity which is related to performance? Our data imply that certain neuroelectric events follow somewhat these expectations, but there are exceptions and at present the data base from the Phase I proof of concept effort reported here is insufficient to rule out other interpretations.

In terms of software development (goal 3 above), we have built upon software developed under a previous project for the Air Force. The scoring routine has been made interactive so that it combines automatic (computer determined) and operator determined saccades and eye blinks. This step was deemed important to the goal of developing a fully automated scoring technique for bioelectric events.

The proposed Phase II is intended to follow logically from the accomplishments of Phase I. If a Phase II is awarded, we plan to develop a bundled hardware and software package which will measure characteristics of an individual's eye movement patterns which predict the task demands (workload) and separate them from other eye movement characteristics which are related to time on task. If such information were available with sufficient precision, it would be possible for the USAF to take the output of an individual's interactions with displayed information and modify the display so the system will better accommodate the individual's needs. Pending success of Phase II, a Phase III, funded by internal research and development capital from Essex, will prepare a prototype tool to be marketed in the private sector.

REFERENCES

Bateman, R. P. An heuristic approach to work analysis. Proceedings of the 23rd Annual Meeting of the Human Factors Society, pp. 554-557, Human Factors Society, Santa Monica, CA, 1979.

Bauer, L. O., Goldstein, R., & Stern, J. A. Effects of information-processing demands on physiological response patterns. Human Factors, 1987, 29(2), 213-234.

Benel, R. A., Coles, M. G. H., Benel, D. C. R. Electrodermal lability and individual differences in simultaneous monitoring and tracking. Proceedings of the 23rd Annual Meeting of the Human Factors Society. Human Factors Society, Santa Monica, 1979.

Chapanis, A., Garner, W. R., Morgan, C. T., & Sanford, F. H. Lectures on men and machines: An introduction to human engineering. Systems Research Laboratory, Baltimore, MD, 1947

Cohen, B., Feldman, M., & Diamond, S. P. Effects of eye movement, brain-stem stimulation, and alertness on transmission through lateral geniculate body of monkey. Journal of Neurophysiology, 1969, 32, 583-594.

Collins, W. E., Crampton, W. E., & Posner, J. B. Effects of mental activity on vestibular nystagmus and the electroencephalogram. Nature, 1961, 190, 194-195.

Collins, W. E., & Posner, J. B. Electroencephalogram alpha activity during mild vestibular stimulation. Nature, 1963, 199, 933-934.

Crabtree, M. S., Bateman, R. P., & Acton, W. H. Benefits of using objective and subjective workload measures. Proceedings of the Human Factors Society 28th Annual Meeting, pp. 950-953, Human Factors Society, Santa Monica, 1984.

Damos, D. Subjective workload and individual differences in information processing abilities. Proceedings of the Fourth Annual Aerospace Behavioral Engineering Technology Conference, pp. 71-74, 1984a.

Damos, D. L. Individual differences in multiple-task performance and subjective estimates of workload. Perceptual and Motor Skills, 1984b, 59, 567-580.

Damos, D. L. Examining the relationship between subjective estimates of workload and individual differences in performance, NASA CR-2341. National Aeronautics and Space Administration, Washington, DC, 1984c

Damos, D., & Smist, T. Individual differences in multi-task response strategies. Proceedings of the Human Factors Society 25th Annual Meeting, pp. 291-295, Human Factors Society, Santa Monica, 1981.

Daroff, R. B., & Hoyt, W. F. In Bach-y-Rita and Collins (Eds.), Symposium on the Control of Eye Movements. Academic Press, New York, 1971.

Eggemeier, F. T. Some current issues in workload assessment. Proceedings of the Human Factors Society 24th Annual Meeting, pp. 669-673. Human Factors Society, Santa Monica, CA, 1980.

Eggemeier, F. T., Shingledecker, C. A., & Crabtree, M. S. Workload measurement in system design and evaluation. Proceedings of the Human Factors Society 29th Annual Meeting, pp. 215-219, Human Factors Society, Santa Monica, CA, 1985.

Frazier, M. L., & Crombie, R. B. (Eds.). Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards Air Force Base, CA, 1982. (DTIC No. AD-A129 333).

Gopher, D., & Donchin, E. Workload--an examination of the concept. In. K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), Handbook of perception and human performance. Wiley, New York, 1986.

Hale, H. B., Williams, Smith, B. N., & Melton, C. E., Jr. Excretion patterns of air traffic controllers. Aerospace Medicine, 1971, 42(2), 127-138.

House Armed Services Committee, Defense Policy Panel Hearings on the downing of Iran Air Flight 655, The effects of human factors on combat performance, October 1988.

Jerison, H. J. Effect of a combination of noise and fatigue on a complex counting task, WADC TR-55-360. Wright Air Development Center, Wright-Patterson AFB, Dayton, OH, 1956.

Kennedy, R. S. A comparison of performance on visual and auditory monitoring tasks. Human Factors, 1971, 13(2), 93-97.

Kennedy, R. S. The relationship between habituation to vestibular stimulation and vigilance: Individual differences and subsidiary problems, NAMRL Monograph No. 20, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1975.

Kennedy, R. S. The relationship between vigilance and eye movements induced by vestibular stimulation. In R. R. Mackie (Ed.), Vigilance: Theory, operational performance, and physiological correlates, pp. 575-602. New York, Plenum.

Kennedy, R. S. Bioelectric indicants of diver's ability to perform useful work. Proceedings of the Undersea Medical Society Workshop, Bethesda, MD, 1978. (AD A060-675)

Kennedy, R. S., & Bittner, A. C., Jr. Development of Performance Evaluation Tests for Environmental Research (PETER): Complex counting. Aviation, Space, and Environmental Medicine, 1980, 51, 142-144.

Kennedy, R. S., May, J. G., Jones, M. B., & Fowlkes, J. E. (1989). Development of saccade length index of taskload for biocybernetic application (Final Tech. Rep, Contract F49620-87-C-0002). Bolling Air Force Base, DC: Air Force Office of Scientific Research.

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988, July). Optimal solutions for complex design problems: Using isoperformance software for human factors trade-offs. Paper presented at the Space Operations Automation and Robotics Workshop: Space Application of Artificial Intelligence, Human Factors, and Robotics, Dayton, OH.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. Eye movements as an index of mental workload, Final Report, Contract F49620-85-C-0121, Air Force Office of Scientific Research, Bolling Air Force Base, DC, 1985.

Merriman, S. C., & Chatelier, P. R. Impact on the Department of Defense Science and Technology Program on training and personnel systems technology. Office of Environmental and Life Sciences, Washington, DC, 1981. (NTIS No. AD A103382)

Orlando Sentinel. (1989, June 6). Stealth jet pilots fatigued (p. A-7).

Schmidt, D. K. A queueing analysis of the air traffic controller's workload. IEEE Transactions on Systems, Man, and Cybernetics, pp. 492-498, SMC-8, 1978.

Schmitt, H. H., & Reid, D. J. Anecdotal information on space adaptation syndrome. NASA/Space Biomedical Research Institute, USRA/Division of Space Biomedicine, July 1985.

Smith, K. U. Cybernetic theory and analysis of learning. In E. A. Bilodeau (Ed.), Acquisition of skill. Academic Press, pp. 425-482, New York, 1966.

Smith, K. U. Cybernetic foundations of physical behavior science. Quest, 1967, 8, 26-89.

Smith, K. U., & Smith, M. F. Cybernetic principles of learning and educational design. Holt, Rinehart, & Winston, New York, 1966.

Stern, J. A., & Skelly, J. J. The eye blink and workload considerations. Proceedings of the Human Factors Society 28th Annual Meeting, 1984, 2, 942-944.

Tice, J. Army program integrates personnel training and weapons design. Defense News, p. 6, January 1986.

Wickens, C. D., Mountford, S. J., & Schreiner, W. Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. Human Factors, 1981, 23, 211-229.

Wiener, N. Cybernetics. Wiley, New York, 1948.

Yules, R. B., Krebs, C. Q., & Gault, F. P. Reticular formation control of vestibular systems. <u>Experimental Neurology</u>, 1966, <u>16</u>, 349-358.